**Machine Learning - Classification**

**SVM Grid Search:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.88 | 0.92 | 0.90 | 123 |
| **1** | 0.86 | 0.79 | 0.82 | 77 |
|  |  |  |  |  |
| **accuracy** |  |  | 0.87 | 200 |
| **macroavg** | 0.87 | 0.86 | 0.86 | 200 |
| **weighted avg** | 0.87 | 0.87 | 0.87 | 200 |

**Random Forest Grid Search:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.61 | 0.72 | 0.66 | 123 |
| **1** | 0.36 | 0.25 | 0.29 | 77 |
|  |  |  |  |  |
| **accuracy** |  |  | 0.54 | 200 |
| **macroavg** | 0.48 | 0.49 | 0.48 | 200 |
| **weighted avg** | 0.51 | 0.54 | 0.52 | 200 |

**Decision Tree Grid Search:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.85 | 0.92 | 0.88 | 123 |
| **1** | 0.85 | 0.74 | 0.79 | 77 |
|  |  |  |  |  |
| **accuracy** |  |  | 0.85 | 200 |
| **macroavg** | 0.85 | 0.83 | 0.84 | 200 |
| **weighted avg** | 0.85 | 0.85 | 0.85 | 200 |

**Good Model here is :**

**SVM Model : Accuracy :** 0.87

**Questions:**

1. How Class is classified?

Ans: Class is classified as Balanced and Imbalanced using **value\_counts()**

1. What is Balanced Class?

Ans: Balanced Class are equally partitioned number of the dataset in the Matrix, which is 50% of the data for each class.

Unbalanced Class are unequally partitioned number of the dataset in the matrix.

1. What is the advantage of balanced dataset?

Ans: There is no partially to learn the dataset. Prediction will be accurate.

1. What is Accuracy?

Ans: How often Model predicts the correct evaluation. percentage/proportion of correct classification of all class to the Total dataset (Data 1 + Data2)

Accuracy = True (Data1) + True (Data2)

True (Data1) + T(Data2) + F(True) + F(Data2) 🡪 Total input of the test set

1. What is Recall?

Ans: Predicts the metrics for each class. How many of the actual positive cases the model was able to identify.

Proportion of correctly classified positive with the all-actual positive datasets

Recall = True (Data1)

Total (data1)

1. Precision:

Ans: Proportion of correctly classified positive with the correctly classified and wrongly classified prediction from one data in dataset.

How many of the actual positive predictions made by the model was correct.

Precision = True (Data1)

True (Data1) + False (Data1)

1. F1- Measure:

F1- Measure is to identify the imbalanced data, whereas it is to identify between Recall and Precision.

IF Recall is high, Precision is low -> We have to identify whether it is good model or not.

F1 = 2 \* (Recall \* Precision)

(Recall + Precision)

1. Macro – Average:

Ans: Average of Recall, Precision and F1Score.

1. Weighted-Average – Sum of product of proportion rate of each class.

(Precision \* Total Data1/Total Set) + (Precision \* Total Data2/Total Set)

(Recall \* Total Data1/Total Set) + (Recall \* Total Data2/Total Set)

(F1 \* Total Data1/Total Set) + (F1 \* Total Data2/Total Set)

1. What is the percentage of correct classification all class to the total input of test set. – Accuracy
2. What is the percentage of correct classification of each class to the total input of test set? - Recall
3. What is the percentage of correct classification of each class to the sum of correctly classified and wrongly classified of each class in the test set? – Precision
4. What is the overall performance of each class? – F1-Measure
5. What if the recall value is high and Precision Value is low. How will you validate your model Performance? - Check the F1-Measure
6. If recall and Precision both are high? – Good Model. Performs good
7. What is the average performance of each class in the input dataset.
8. What is the sum of product of proportion rate of each class? - Weighted Average